From:	Shulman, Stu
To:	Bowman, Randal
Cc:	Cash, Marcia
Subject:	Re: Stu"s Email
Date:	Sunday, May 28, 2017 4:27:00 PM
Attachments:	image.png

On Sun, May 28, 2017 at 5:14 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

> I'm not being critical, and "not available" is apparently another example of our using terms differently. I don't recall any comments about computational delays, although that is quite reasonable once explained.
>
> Will have draft codes momentarily for both your review.
>
> On Sun, May 28, 2017 at 5:08 PM, Shulman, Stu <stu@texifter.com> wrote:
>> I believe both Marcia and I indicated during the that once the data is uploaded the groupings take time to complete. The computational process required indexing every 2-word combination in the collection of 117,000 documents and comparing them to every other 2-word combination in all 117,000 documents is a huge. There is a lot of high-powered math going on behind the scenes.
>>
>> There are no features that are not "available" rather, as noted, the computing time is a part of the feature in operation.
>>
>>
>> On Sun, May 28, 2017 at 4:57 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:
>>
>>> yes to meeting tonight, at your convenience. I will send my thoughts on codes shortly.
>>>
>>> I think a big part of the problem here is my lack of familiarity with the basic wording associated with the system, and with the system. I had presumed that once the comments were loaded, everything would immediately be up and running. On our call, I will set out exactly what I would like to accomplish tonight and/or tomorrow, and you can let me know when those features are likely to be available, along with whatever else I appear to need to know. I have other work related to the project that I can do in meantime.
>>>
>>>
>>> On Sun, May 28, 2017 at 4:47 PM, Shulman, Stu <stu@texifter.com> wrote:
>>>> Randy,
>>>>
>>>> We are moving as fast as possible. The reasons I thought you were exploring the comments and software were:
>>>>
>>>> - to test your hunches,
>>>> - to learn the software basics, and

- to become more familiar with the comments

I did not understand from our previous discussions that you would be using this weekend to do the actual comment review. If that is the case, send me your final coding scheme, fully fleshed out, and when the Bears Ears portion is done deduplicating I can set it up and you can start.

What I thought we were doing is having a meeting tonight to discuss your options and review the work you had done exploring the comments today. Did you still want a meeting tonight, and if so, at which time?

Stu

On Sun, May 28, 2017 at 3:49 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

> points taken re coding. However, we are in a time crunch for the Bears Ears comments, which must be finished by June 7. I really wanted to use the time this weekend - apparently tomorrow now, if then - to clear out all of the duplicate and near-duplicate comments by coding them into their appropriate categories, and then if at all possible start applying the word sets I drafted to see if those create additional large groupings as I anticipate. You have my thoughts on types of codes, and I will defer to your experience if using 1 and 2 or 1 and 9 for opposing and supporting comments is preferable. Since the opposing comments are the overwhelming majority, they should be "1" in any case.
>
> However, if things can't go that fast I understand - the contracting delays were unavoidable - but there is no point in my just reading comments.
>
> On Sun, May 28, 2017 at 3:33 PM, Shulman, Stu <stu@texifter.com> wrote:
>> Randy,
>>
>> Please see me previous explanation about the groups. Duplicate detection is not going to have the effect of filtering out non-Bears Ears comments. I'm not sure how else to explain that.
>>
>> There is, however, a bucket with 59,559 comments that mention Bears Ears:
>>
>> http://doi.discovertext.com/app/bucketDetails.aspx?nloc=0000 00040000000000000012000000000000000000000000
>>
>> I am in the procession of converting that bucket into a new archive. Then I will de-duplicate the comments that are just about Bears Ears.
>>
>> With respect to creating code sets, I really think you should leave that to me. I am the data manager and the experienced software user. There is nothing wrong with you exploring the data and refining your thinking about how best to proceed, but as I understood it, the DOI hired me to do the management part precisely because I am not a novice user of the software.

If you could, please consider the work you do this weekend exploratory, and the meeting tonight as the best chance to efficiently set up a plan to get through the data. I do not think it makes sense for you to start coding on your own when you have no experience with the software or project management on this platform.

Stu

On Sun, May 28, 2017 at 3:23 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

> I tried to create codes using the instructions in the link above, but cannot get the screens the directions show to appear with the real data - I presume because it is still loading. Please send either an email or a message on the system when everything is loaded and there are only Bears Ears comments available for review, and I'll start again. Will check every 30 minutes or so. Also please let me know when you want to talk.
>
> And examples of non-Bears Ears comments now - groups 2 and 3, approx 2,000 comments, do not include the words Bears Ears in the comments.
>
> On Sun, May 28, 2017 at 2:31 PM, Shulman, Stu <stu@texifter.com> wrote:
>> Randy,
>>
>> Once the final documents are in we will be redoing the groups. I think you should hold off doing more than exploring the data until the ingest is complete.
>>
>> All of the features of the system are documented here:
>>
>> https://texifter.zendesk.com/hc/en-us
>>
>> Mark indicates it is the OCR process that pulls text from scanned documents that is slowing the process to a crawl on the last batch.
>>
>> Stu
>>
>>
>>
>> On Sun, May 28, 2017 at 2:25 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:
>>
>>> Thank you. I am reading duplicate groups now, but have forgotten how to create codes and apply them to the groups, so will keep reading until you have time to send me directions.
>>>
>>> On Sun, May 28, 2017 at 2:22 PM, Shulman, Stu <stu@texifter.com> wrote:
>>>> Randy,
>>>>
>>>> There are a total of 29 .zip archives. 27 have completed, but the last two

are still processing. This may have something to do with the presence of large attachments.

I'll keep monitoring the process.

Stu

On Sun, May 28, 2017 at 12:33 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

> change in plans - it started raining lightly while I was typing, so no lawnmowing today. Will check every half hour or so to see if full Bears Ears data set is posted starting about 2.
>
> On Sun, May 28, 2017 at 9:14 AM, Shulman, Stu <stu@texifter.com> wrote:
>> Hold the phone.
>>
>> Mark (the chief engineer) just told me there are still more comments ingesting and that those duplicate detection results were from a false start yesterday,
>>
>> So, you can definitely experiment with what is showing, but we are still waiting for the final data to complete indexing and we will need to re-do the duplicate detection and clustering.
>>
>> Stu
>>
>> On Sun, May 28, 2017 at 8:58 AM, Shulman, Stu <stu@texifter.com> wrote:
>>> Randy,
>>>
>>> There were 116,417 comments altogether so the ingest and indexing takes time to run; it had to go overnight to find the duplicates. I have shared the project with you and you have full access to the data. A few notes:
>>>
>>> - Your level of access includes the ability to delete all the data and the project.
>>> - The duplicate detection is complete.
>>> - The near-duplicate clustering is still running
>>> - A search for "Bears Ears" in the raw data shows 59,225 results
>>> - There is a "bucket" with all 59,225 you can experiment with while the clustering completes
>>> - The biggest duplicate group has 13,372 identical comments "As a supporter of bird conservation and our public lands, I strongly urge you to protect..."

I am around tonight if you want to have a web meeting. Probably 7 PM or 9 PM would be best.

Stu

On Sat, May 27, 2017 at 6:01 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

> I have registered and accessed the dashboard with no problems. However, I cannot see how to access the comments - or are they not posted yet? I will check back around 9 pm with the hope of doing a little work tonight and much more tomorrow
>
> On Fri, May 26, 2017 at 4:17 PM, Shulman, Stu <stu@texifter.com> wrote:
>
>> --
>> Dr. Stuart W. Shulman
>> Founder and CEO, Texifter
>> Cell: 413-992-8513
>> LinkedIn: http://www.linkedin.com/in/stuartwshulman
>
> --
> Dr. Stuart W. Shulman
> Founder and CEO, Texifter
> Cell: 413-992-8513
> LinkedIn: http://www.linkedin.com/in/stuartwshulman

--
Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: 413-992-8513
LinkedIn: http://www.linkedin.com/in/stuartwshulman

--
Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: 413-992-8513

LinkedIn: http://www.linkedin.com/in/stuartwshulman

acres act america american americans antiquities archaeological atlantic attack
believe belongs benefited bird birds boundaries buttes canyons care change changes
comment communities consequences conservation country created critical cultural
current currently depend designated designation designations dont eagle
economic economically ecosystem enjoy executive existing federal fifteen future
generations golden grand habitat hanford harmful help helped heritage historic

history iconic including incredible indispensable input land lands landscapes
law legacy live local magnificent maintain management mistake mountainsdesert
nation nations native natural northeast objects organ outdoor pacific
papahanaumokuakea parks partieshave past peaks people please preserve
president presidents presidentsof pristine process protect protected protecting

protection protections public puffins reach recognized recreation reject
remain rescinding resources reversing review rich rocky roosevelt roosevelts
safeguard seamounts shrubsteppe signed sites size staircase staircaseescalante status
story strongly subsequent support supporter teddy thank time tragic trails
treasures tribes trump twenty underwater uphold urge utah valuable value visit
vital washington waters wild wildlife

generated by **DiscoverText**